



Search for Standard Model Higgs Boson Production in Association With a W Boson Using Matrix Element and Boosted Decision Tree Techniques With 2.7 fb^{-1} of Data

The CDF Collaboration
URL <http://www-cdf.fnal.gov>
(Dated: January 28, 2009)

We present a search for Standard Model Higgs boson production in association with a W boson using 2.7 fb^{-1} of CDF II data collected between 2002 and 2008. This search is performed using two complementary techniques: a matrix element technique is used to calculate event probability densities for the signal and background hypothesis and a multivariate technique, based on boosted decision tree, is used to combine the event probability densities with kinematic variables to build a final discriminant distribution which is fitted to the data using a binned likelihood approach.

We observe no evidence for a Higgs boson signal and set 95% confidence level upper limits on the WH production cross section times the branching ratio of the Higgs boson to decay to $b\bar{b}$ pairs of $\sigma(p\bar{p} \rightarrow WH) \times BR(H \rightarrow b\bar{b})/SM < 3.34$ to 84.1 for Higgs boson masses between $m_H = 100 \text{ GeV}/c^2$ and $m_H = 150 \text{ GeV}/c^2$. The expected (median) limit estimated in pseudo-experiments is: $\sigma(p\bar{p} \rightarrow WH) \times BR(H \rightarrow b\bar{b})/SM < 3.80$ to 48.7 at 95% C.L.

Preliminary Results for Summer 2008 Conferences

Finding evidence for Higgs boson production in association with a W boson is extremely challenging since at best it is rarely produced ($\sigma_{WH} \sim 0.1$ pb) in comparison with other processes with the same final state like $W + b\bar{b}$ and top. The signal to background ratio of the analysis is expected to be tiny, typically on the order of $S/B \sim 1/100$. In this note we describe two complementary analysis techniques combined in order to achieve better discrimination of signal and background events of the combination.

This note describes the search for WH production using a matrix element analysis combined with boosted decision trees. The implementation of these methods is very similar to the single-top search described in detail here [1, 2]. We combine both techniques using the discriminant from the matrix element analysis as an input variable of the Boosted Decision analysis.

DATA SAMPLE & EVENT SELECTION

The candidate events for this analysis are selected by requiring a $W + 2$ jet event topology where the W decays leptonically, $W \rightarrow e\nu_e$ or $W \rightarrow \mu\nu_\mu$. We require events to contain an isolated electron or muon with offline E_T or $p_T > 20$ GeV and Missing Transverse Energy, $\cancel{E}_T > 20$ GeV (25 GeV for forward electrons). Jets are clustered with a cone size of $\Delta R < 0.4$ and are required to have $E_T > 20$ GeV after correcting the jets for instrumental effects and have $|\eta_{detector}| < 2.0$. One or both of the two jets should be identified as a b -jet using the secondary vertex tag requirement. The secondary vertex tagging algorithm identifies tracks associated with the jet originating from a vertex displaced from the primary vertex indicative of decay particles from relatively long lived B mesons. We veto dilepton, Z -boson, leptons from photon conversion and QCD multi-jet events.

BACKGROUND ESTIMATE

Estimating the background contribution after applying the event selection to the WH candidate sample is an elaborate process. The backgrounds surviving the selection are $t\bar{t}$, s-channel and t-channel single top, $W +$ heavy-flavor jets, i.e. $W + b\bar{b}$, $W + c\bar{c}$ and $W + c$, and diboson events WW , WZ , and ZZ . Instrumental backgrounds originate from mis-tagged $W +$ jets events (W events with light-flavor jets, i.e. with u , d , s -quark and gluon content, misidentified as heavy-flavor jets) and from non- $W +$ jets events (multi-jet events where one jet is erroneously identified as a lepton). We determine the $W +$ jets normalization from the data and estimate the fraction of the candidate events with heavy-flavor jets using ALPGEN Monte Carlo samples [3]. The heavy-flavor fractions were calibrated in the b -tagged $W + 1$ jet sample using data distributions which are sensitive to distinguish light-flavor from heavy-flavor jets, e.g. the mass of the secondary-vertex and, more sophisticated, the output of the Neural Network jet-flavor separator. Based on these studies, the heavy flavor content was corrected by a factor $K_{HF} = 1.4 \pm 0.4$. The probability that a $W +$ light-flavor jet is mis-tagged is parameterized using large statistics generic multi-jet data. The instrumental background contribution from non- W events is estimated using side-band data with low missing transverse energy, devoid of any signal, and we subsequently extrapolate the contribution into the signal region with large missing transverse energy, $\cancel{E}_T > 20$ GeV (25 GeV for forward electrons).

For the MC based background estimations, we have used the theoretical cross sections shown in Table I. NLO cross section calculations exist for diboson and $t\bar{t}$ production, thereby making the estimation of their contribution a relatively straightforward process. We have added the expected contribution from WH signal using the SM cross sections and branching ratios shown in Table II. The expected signal and background yield in the $W + 2$ jet sample is shown in Table III.

Process	Theoretical Cross Section
s-channel	0.884 ± 0.11
t-channel	1.980 ± 0.25
WW	12.4 ± 0.25
WZ	3.96 ± 0.06
ZZ	1.58 ± 0.05
$t\bar{t}$	6.7 ± 0.8
Z +jets	787.4 ± 85.0

TABLE I: Theoretical cross sections used for the MC based background estimation.

Higgs Mass (GeV/c ²)	BR($H \rightarrow b\bar{b}$)	σ (pb)	$\sigma \times \text{BR}(H \rightarrow b\bar{b})$ (pb)
100	0.812	0.286	0.232
105	0.796	0.253	0.201
110	0.770	0.219	0.169
115	0.732	0.186	0.136
120	0.679	0.153	0.104
125	0.610	0.136	0.083
130	0.527	0.120	0.063
135	0.436	0.103	0.045
140	0.344	0.086	0.030
145	0.256	0.078	0.020
150	0.176	0.070	0.012

TABLE II: SM branching ratios ($H \rightarrow b\bar{b}$) and cross sections for all Higgs masses.

Process	1 tag	2 tags
All Pretag Cands.	50644.0 \pm 0.0	57174.0 \pm 0.0
WW	56.2 \pm 6.2	0.4 \pm 0.1
WZ	23.0 \pm 1.7	4.8 \pm 0.5
ZZ	0.8 \pm 0.1	0.2 \pm 0.0
TopLJ	121.3 \pm 17.1	23.8 \pm 3.9
TopDil	48.8 \pm 6.8	14.1 \pm 2.3
STopT	64.0 \pm 9.3	1.8 \pm 0.3
STopS	40.6 \pm 5.7	12.8 \pm 2.1
Z+jets	37.4 \pm 5.5	2.1 \pm 0.3
Total MC	392.0 \pm 35.0	59.9 \pm 7.5
Wbb	538.7 \pm 162.5	70.3 \pm 22.5
Wcc/Wc	489.1 \pm 150.9	6.8 \pm 2.3
Total HF	1027.8 \pm 312.3	77.1 \pm 24.7
Mistags	458.0 \pm 57.9	2.2 \pm 0.6
Non-W	135.5 \pm 54.2	9.0 \pm 3.6
Total Prediction	2013.3 \pm 324.1	148.2 \pm 26.1
WH100	9.5 \pm 0.8	2.9 \pm 0.3
WH105	8.6 \pm 0.7	2.7 \pm 0.3
WH110	7.6 \pm 0.6	2.4 \pm 0.3
WH115	6.3 \pm 0.5	2.0 \pm 0.2
WH120	4.9 \pm 0.4	1.6 \pm 0.2
WH125	4.0 \pm 0.3	1.3 \pm 0.2
WH130	3.1 \pm 0.3	1.0 \pm 0.1
WH135	2.3 \pm 0.2	0.7 \pm 0.1
WH140	1.5 \pm 0.1	0.5 \pm 0.1
WH145	1.0 \pm 0.1	0.3 \pm 0.0
WH150	0.7 \pm 0.1	0.2 \pm 0.0
Observed	1998.0 \pm 0.0	156.0 \pm 0.0

TABLE III: Number of expected signal and background events, in the 2 jet bin, in 2.7 fb⁻¹ of CDF data, passing all the event selection requirements.

ANALYSIS METHOD

Matrix Element Method

The matrix element analysis relies on the evaluation of event probability densities for signal and background processes based on the Standard Model differential cross-section calculation.

In general a differential cross-section is given by [4]:

$$d\sigma = \frac{(2\pi)^4 |M|^2}{4\sqrt{(q_1 \cdot q_2)^2 - m_{q_1}^2 m_{q_2}^2}} d\Phi_n(q_1 + q_2; p_1, \dots, p_n) \quad (1)$$

where $|M|$ is the Lorentz invariant matrix element; q_1 , q_2 and m_{q_1} , m_{q_2} are the four momenta and masses of the incident particles, and $d\Phi_n$ is the n -body phase space given by [4]:

$$d\Phi_n(q_1 + q_2; p_1, \dots, p_n) = \delta^4(q_1 + q_2 - \sum_{i=1}^n p_i) \prod_{i=1}^n \frac{d^3 p_i}{(2\pi)^3 2E_i} \quad (2)$$

The CDF detector would be ‘ideal’ if we could measure all four momenta of the initial and final state particles perfectly. In this case we could use this formula without modification and normalize it to the total cross section to define the event probability:

$$P_{evt} \sim \frac{d\sigma}{\sigma}$$

However, several effects have to be considered: (1) the initial state interaction is initiated by partons inside the proton and antiproton, (2) neutrinos in the final state are not identified directly, and (3) the energy resolution of the detector can not be ignored. To address the first point, the differential cross section is folded over the parton distribution functions. To address the second and third points, we integrate over all particle momenta which we do not measure (e.g. p_z of the neutrino), or do not measure very well, due to resolution effects (e.g. jet energies). The integration reflects the fact that we want to sum over all possible particle variables (y) leading to the observed set of variables (x) measured with the CDF detector. The mapping between the particle variables (y) and the measured variables (x) is established with the transfer function, $W(y, x)$. After incorporating the effects mentioned above, the event probability takes the form:

$$P(x) = \frac{1}{\sigma} \int d\sigma(y) dq_1 dq_2 f(y_1) f(y_2) W(y, x) \quad (3)$$

where $d\sigma(y)$ is the differential cross section in terms of the particle variables; $f(y_i)$ are the PDFs, with y_i being the fraction of the proton momentum carried by the parton ($y_i = E_{q_i}/E_{beam}$); and $W(y, x)$ is the transfer function. Substituting Equation 1 and 2 into Equation 3, and considering a final state with four particles ($n=4$), the event probability becomes:

$$P(x) = \frac{1}{\sigma} \int 2\pi^4 |M|^2 \frac{f(y_1)}{|E_{q_1}|} \frac{f(y_2)}{|E_{q_2}|} W(y, x) d\Phi_4 dE_{q_1} dE_{q_2} \quad (4)$$

where the masses and transverse momenta of the initial partons are neglected (i.e. $\sqrt{(q_1 \cdot q_2)^2 - m_{q_1}^2 m_{q_2}^2} \simeq 2E_{q_1} E_{q_2}$).

In this analysis, we calculate event probability densities for the WH signal, as well as for the s-channel and t-channel single top, $t\bar{t}$, $Wb\bar{b}$, $Wc\bar{c}$, Mistags (Wgg) and diboson background processes. The Wcc and non- W events are assumed to be represented fairly well by the background probability density.

We calculate the matrix element ($|M|^2$) for the event probability density at leading order perturbation theory by using the HELAS (HELicity Amplitude Subroutines for Feynman Diagram Evaluations) package. The correct subroutines for a given process are automatically generated by the MadGraph program. We use different subroutines for calculating event probabilities for the WH signal and s-channel, t-channel, $Wb\bar{b}$, $Wc\bar{c}$, Wgg , diboson and $t\bar{t}$ background hypotheses.

Transfer Function: $\mathbf{TF}(\mathbf{E}_p, \mathbf{E}_j)$

The transfer function, $W(y, x)$, provides the probability of measuring the set of observable variables (x) that correspond to the set of production variables (y). The set (y) represents all final state particle momenta at the particle level, while the set (x) represents the measured momenta (of the corresponding object) with the CDF detector. In the case of well-measured objects, $W(y, x)$ is taken as a δ -function (i.e. the measured momenta are used in the differential cross section calculation). When the detector resolution cannot be ignored, $W(y, x)$ is taken as a Gaussian-type

function. For unmeasured quantities, like the momenta of the neutrino, the transfer function is unity (the transverse momenta of the neutrino, however, can be inferred from energy and momentum conservation).

The transfer between parton and jet energies is determined by the transfer function $W_{jet}(E_{parton}, E_{jet})$. The jet energy corrections correct the energies of jets in a way that the means of the corrected jet energies and the original parton energies are equal. Such corrections, however, do not account for the shape of the difference in energies: the shape of the $\delta_E = (E_{parton} - E_{jet})$ distribution. We parameterize the light-jets, gluons and b and c jets transfer functions in a different way due to the different kinematics. For light jets, we determine the parameters of the transfer function $W_{jet}(E_{parton}, E_{jet})$ using the light jets in the t -channel single top sample. These transfer functions are also applied to the background probabilities with light-jets.

We also implement transfer functions obtained from gluons which are applied to all the probabilities with gluons in the final state (these transfer functions are derived from the Wc MC samples).

In order to better reproduce the real parton energy for b and c jets, we train a Neural Network (NN) with the Stuttgart Neural Network Simulator (SNNS) using the ROOT interface ROOTSNNS v3.0. The NN is trained using only WH signal events of several Higgs masses (from 100 to 150 GeV/c² in 5 GeV/c² steps).

For the training we use 8 input variables related to the jet kinematics: the total energy of the jet corrected at level 5 (E_j), the sum of the transverse momentum of the tracks in the jet (SumE) [?], the transverse momentum of the jet (p_T), the ϕ and η of the jet, the fraction of energy deposited for the jet in the electromagnetic calorimeter (EMF), the raw (measured) energy of the jet (RawE_j), and the energy of the jets with cone size 0.7 (Ejcone7) [?].

The goal of using a NN to reproduce the parton energy is to use the output of the NN (instead of the jet energy, E_j) as input of the transfer function. Figure 1 shows the difference between the parton energy and the jet energy and the NN output. It is clear that the NN output is closer to the parton energy than the level 5 corrected jet energies and that the distribution is also narrower. Therefore, since the NN output provides a better jet resolution, using it as input of the transfer function should help to improve the performance of the transfer function.

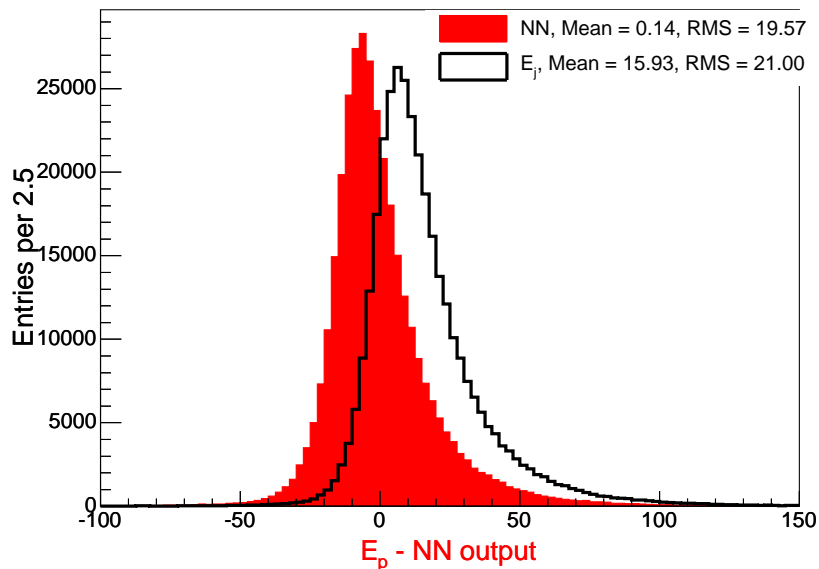


FIG. 1: Difference between the parton energy and the jet energy (empty histogram) and the NN output (red solid histogram).

Event Probability Discriminant

The event probability density makes use of all measured quantities to specify each event. This should provide good discrimination between signal and background. It uses both possible jet combinations in the event so that the right jet-parton association is always included.

We use the event probability densities as ingredients to build an event probability discriminant, i.e. a distribution which separates signal from background which we can use to fit the data. Perhaps the most intuitive discriminant is the ratio of signal over signal + background probability, $EPD = P_s/(P_s + P_b)$. This discriminant is close to zero for ratios dominated by P_b and close to unity for ratios dominated by P_s . The expression 5 is the event probability

discriminant we use in this analysis. We introduce extra non-kinematic information by using the output neural network b-tagger (b) which assigns a probability for each b-tagged jet ($0 < b < 1$).

$$EPD = \frac{b \cdot P_{WH}}{b(P_{WH} + P_{singletop} + P_{Wbb} + P_{t\bar{t}}) + (1-b)(P_{Wc\bar{c}} + P_{Wcj} + P_{Mistag} + P_{diboson})} \quad (5)$$

Boosted Decision Trees Method

In order to search for Higgs production we develop a multivariate technique based on Boosted Decision Trees (BDT). To build the BDTs we make use of the ROOT-integrated package TMVA [5]. This technique has already been applied at CDF for the single top search.

A decision tree is a binary tree structured classifier like the one sketched in Fig. 2. Repeated left/right (yes/no) decisions are performed on a single variable at a time until some stop criterion is reached. Like this the phase space is split into regions that are eventually classified as signal or background, depending on which makes up the majority of training events that end up in the final leaf nodes. The boosting of a decision tree (BDT) represents an extension to a single decision tree. Several decision trees (a forest), derived from the same training sample by reweighting events, are combined to form a classifier which is given by a (weighted) majority vote of the individual decision trees. This process, called boosting, stabilizes the response of the decision trees with respect to fluctuations in the training sample.

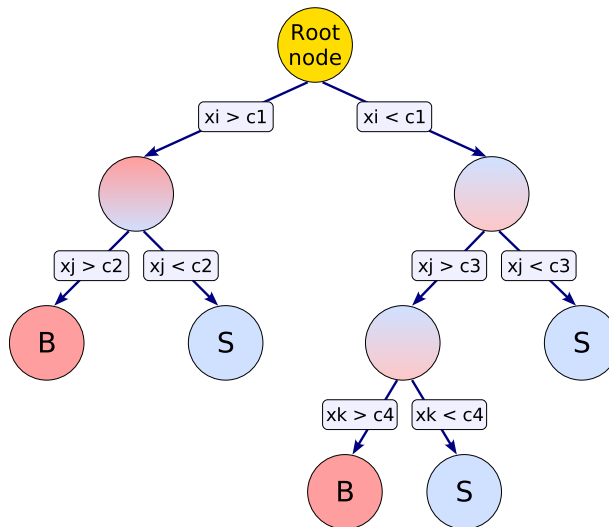


FIG. 2: Schematic view of a decision tree. Starting from the root node, a sequence of binary splits using the discriminating variables x_i is performed. Each split uses the variable that at this node gives the best separation between signal and background when being cut on. The same variable may thus be used at several nodes, while others might not be used at all. The leaf nodes at the bottom end of the tree are labeled S for signal and B for background depending on the majority of events that end up in the respective nodes.

Using boosted decision trees many kinematic or event shape input variables are combined into a single output variable with powerful discrimination between signal and background. In the search for SM Higgs production, two different boosted decision trees are trained for each Higgs mass point and in different b-tag bins:

- 2 jets, 1 b-tag
- 2 jets, 2 b-tags

The level of agreement between data and Monte Carlo simulation is checked for all the input variables in the two signal regions as well as in different control regions.

In addition to the validation of the input variables, the output of the two trained BDTs are validated in the control region of 2 jets with no b-tags ($W + \text{light flavor dominant}$). The outputs of the BDT optimized for the W+2jets-1tag and for the W+2jets-2tags signal regions and evaluated in the control region are shown in Fig. 3.

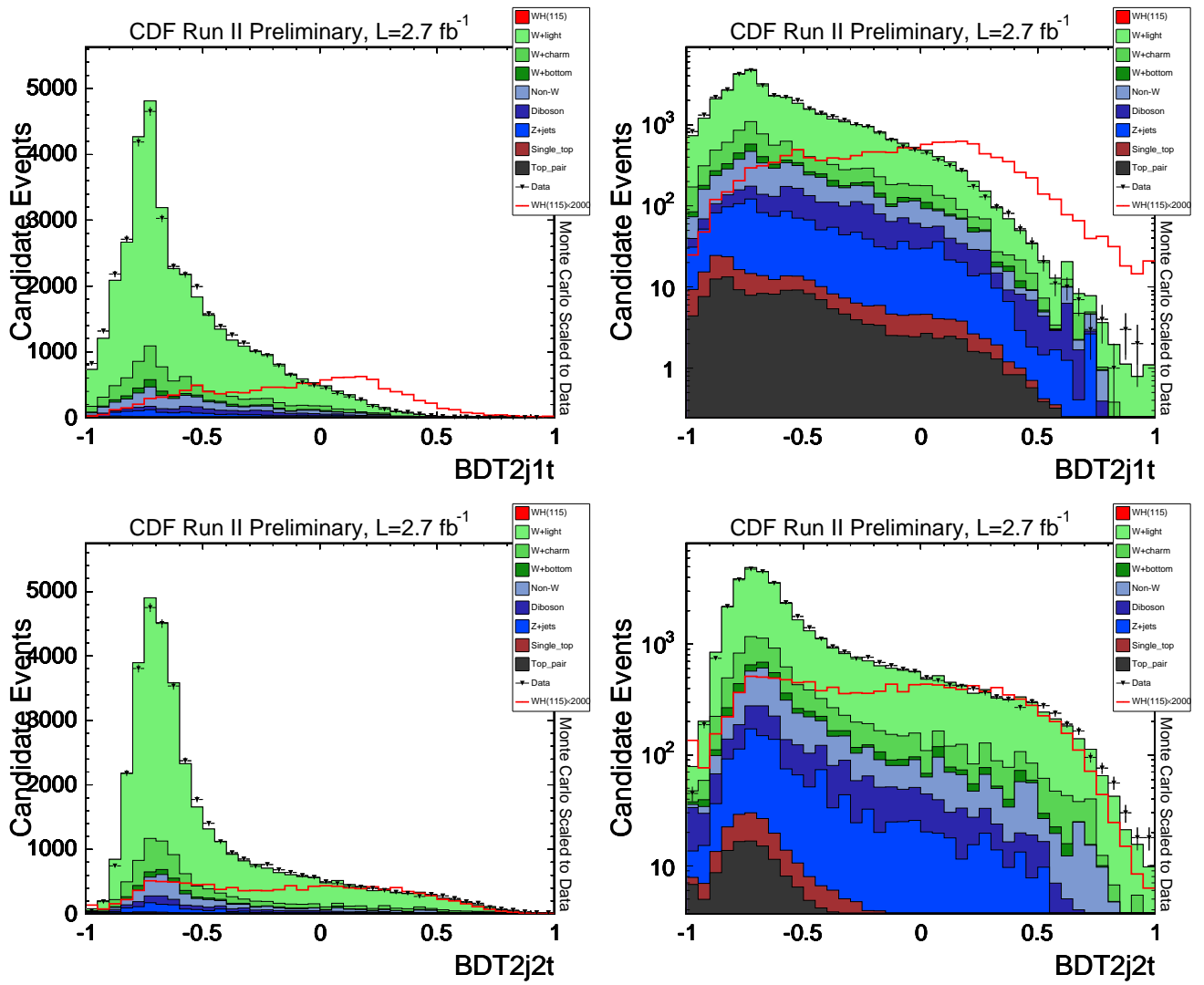


FIG. 3: Distributions of the BDT outputs in the control region with no b-tags ($m_H = 115 \text{ GeV}/c^2$). Left (right) in linear (log) scale, top (bottom) BDT optimized for the W+2jets-1tag (W+2jets-2tags) signal region.

Decision trees are insensitive to the inclusion of poorly discriminating input variables. While for artificial neural networks it is typically more difficult to deal with such additional variables, the decision tree training algorithm will basically ignore non discriminating variables as for each node splitting only the best discriminating variable is used.

The 21 variables used for the training of the BDTs are:

- the event probability discriminant based on matrix element probabilities
- the ratios between the signal event probability and each one of the background event probabilities
- the invariant mass of the di-jet system m_{j1j2}
- the E_T of both jets
- the $\Delta\phi$ between the jets and the \cancel{E}_T
- the $\Delta\phi$ between the the lepton and the \cancel{E}_T
- the p_T and the η of the lepton
- the scalar sum of the transverse energies $H_T = \sum_{jets} E_T + p_T + \cancel{E}_T$

- the cosine of the angle between the lepton and jets
- the transverse mass of the W boson $m_T(W)$
- the KIT NN flavor separator
- the missing transverse energy in the event, \cancel{E}_T

We use the output of the BDT trained in the two channels as the discriminant for a WH search.

THE LIKELIHOOD FUNCTION

The likelihood function, \mathcal{L} , is a function of the unknown Poisson means for signal and background and is defined such that it expresses the joint probability of observing the N data events at their respective values of the BDT output. The values of the Poisson means at which \mathcal{L} achieves its maximum, corresponds to the most probable estimate for the true signal and background content in the data sample.

We perform a binned likelihood fit to the BDT output. To make it easier to compare the different fit parameters, we define the fit parameter as $\beta_j = \sigma_j^{Fit} / \sigma_j^{SM}$ where β_j is unity when the fit result corresponds to the expected number of events obtained from the independent signal/background estimate:

$$\mathcal{L} = \prod_{j=2}^5 G_j(\beta_j; \sigma_j) \prod_{k=1}^B \frac{e^{-\mu_k} \cdot \mu_k^{n_k}}{n_k!} \quad (6)$$

The Gaussian constraints to the backgrounds are given by:

$$G_j(\beta_j; \sigma_j) = \frac{1}{\sqrt{2\pi \cdot \sigma_j^2}} \exp \left[-\frac{1}{2} \cdot \left(\frac{\beta_j - 1.0}{\sigma_j} \right)^2 \right] \quad (7)$$

$$\mu_k = \beta_{\text{single top}} \cdot T_{jk} + \beta_{W+\text{bottom jets}} \cdot T_{jk} + \beta_{W+\text{charm jets}} \cdot T_{jk} + \beta_{\text{mistags}} \cdot T_{jk} + \beta_{t\bar{t}} \cdot T_{jk} \quad (8)$$

The index k runs over the bins of the fitted histogram. The template histograms are normalized to the predicted number of events as shown in Table III. This means, $\sum_{k=1}^B T_{jk} = N_j^{pred}$.

In addition, the prediction in each bin needs an additional Gaussian uncertainty due to the limitations of Monte Carlo statistics. Each bin is allowed to fluctuate according to the total uncertainty in that bin, which is the sum in quadrature of the weight of each event. This prevents us from overestimating our sensitivity due to a fluctuation in Monte Carlo.

SYSTEMATIC UNCERTAINTIES

Systematic uncertainties can bias the outcome of this analysis and have to be incorporated into the result. We address systematic uncertainty from several different sources: (1) jet energy scale, (2) initial state radiation, (3) final state radiation, (4) parton distribution functions, (5) luminosity and (6) b tagging SF as estimated from the most discrepant shapes in the control variables.

Systematic uncertainties can influence both the expected event yield (normalization) and the shape of the discriminant distribution.

Normalization uncertainties are estimated by recalculating the acceptance using Monte Carlo samples altered due to a specific systematic effect. The WH normalization uncertainty is the difference between the systematically shifted acceptance and the default one and are shown in Table IV.

The effect of the uncertainty in the jet energy scale is evaluated by applying jet-energy corrections that describe $\pm 1\sigma$ variations to the default correction factor. Systematic uncertainties due to the modeling of ISR and FSR are obtained from dedicated Monte Carlo samples where the strength of ISR/FSR was increased and decreased in the parton showering to represent $\pm 1\sigma$ variations.

Channel	Lepton ID	Luminosity	b -tagging SF	ISR/FSR + PDF	JES
single-tag	$\sim 2\%$	6%	3.5%	3.1%	2.0%
double-tag	$\sim 2\%$	6%	8.4%-9.1%	4.3%-5.6%	2.0%

TABLE IV: Rate systematic uncertainties for each channel.

The effect of the b -tag scale factor and luminosity uncertainty is determined from the background estimate (for the signal template only; the background templates have these numbers included in their Gaussian constraints).

For all backgrounds the normalization uncertainties are represented by the uncertainty on the predicted number of background events and are incorporated in the analysis as Gaussian constraints $G(\beta_j|1, \Delta_j)$ in the likelihood function:

$$\mathcal{L}(\beta_1, \dots, \beta_5; \delta_1, \dots, \delta_{10}) = \underbrace{\prod_{k=1}^B \frac{e^{-\mu_k} \cdot \mu_k^{n_k}}{n_k!}}_{\text{Poisson term}} \cdot \underbrace{\prod_{j=2}^5 G(\beta_j|1, \Delta_j)}_{\text{Gauss constraints}} \cdot \underbrace{\prod_{i=1}^{12} G(\delta_i, 0, 1)}_{\text{Systematics}} \quad (9)$$

$$\text{where, } \mu_k = \sum_{j=1}^5 \beta_j \cdot \underbrace{\left\{ \prod_{i=1}^{12} [1 + |\delta_i| \cdot (\epsilon_{ji+} H(\delta_i) + \epsilon_{ji-} H(-\delta_i))] \right\}}_{\text{Normalization Uncertainty}} \quad (10)$$

$$\cdot \underbrace{\alpha_{jk}}_{\text{Shape P.}} \cdot \underbrace{\left\{ \prod_{i=1}^{12} (1 + |\delta_i| \cdot (\kappa_{jik+} H(\delta_i) + \kappa_{jik-} H(-\delta_i))) \right\}}_{\text{Shape Uncertainty}} \quad (11)$$

The systematic normalizations are incorporated into the likelihood as nuisance parameters, conforming with a fully Bayesian treatment [6]. We take the correlation between normalization for a given source into account [7]. The relative strength of a systematic effect due to the source i is parameterized by the nuisance parameter δ_i in the likelihood function, constrained to a unit-width Gaussian (last term in Equation 9). The $\pm 1\sigma$ changes in the normalization of process j due to the i^{th} source of systematic uncertainty are denoted by ϵ_{ji+} and ϵ_{ji-} (see Equation part 10). The $\pm 1\sigma$ changes in bin k of the templates for process j due to the i^{th} source of systematic uncertainty are quantified by κ_{jik+} . $H(\delta_i)$ represents the Heaviside function, defined as $H(\delta_i) = 1$ for $\delta_i > 0$ and $H(\delta_i) = 0$ for $\delta_i < 0$. The Heaviside function is used to separate positive and negative systematic shifts (for which we have different normalization). The variable δ_i appears in both the term for the normalization (Equation 10), which is how correlations between both effects are taken into account. We marginalizing the likelihood function by integrating $\mathcal{L}(\beta_1, \dots, \beta_N, \delta_1, \dots, \delta_S)$ over all nuisance parameters for many possible values of the WH cross-section $\beta_1 = \beta_{WH}$. The resulting reduced likelihood $\mathcal{L}(\beta_{WH})$ is a function of the WH cross-section β_{WH} only. We use the MCLIMIT package for our statistical treatment [8].

The event detection efficiency includes uncertainties on the lepton ID, trigger efficiencies and b -tagging scale-factors. The uncertainties on the data derived backgrounds (W +bottom, W +charm, mistags and non- W) are taken from the event yield in Table III.

RESULTS WITH CDF II DATA

We apply the analysis to 2.7 fb^{-1} of CDF Run II data. We compare the BDT output distribution, for a Higgs mass of $115 \text{ GeV}/c^2$, of our candidate events with the sum of predicted WH signal and background distributions as shown in Fig. 4.

In order to extract the most probable WH signal content in the data we perform the maximum likelihood method described before. We perform marginalization using the likelihood function of Equation 9 with all systematic uncertainties included in the likelihood function. The posterior p.d.f is obtained by using Bayes' theorem:

$$p(\beta_1|data) = \frac{\mathcal{L}^*(data|\beta_{WH})\pi(\beta_{WH})}{\int \mathcal{L}^*(data|\beta'_{WH})\pi(\beta'_{WH})d\beta'_{WH}}$$

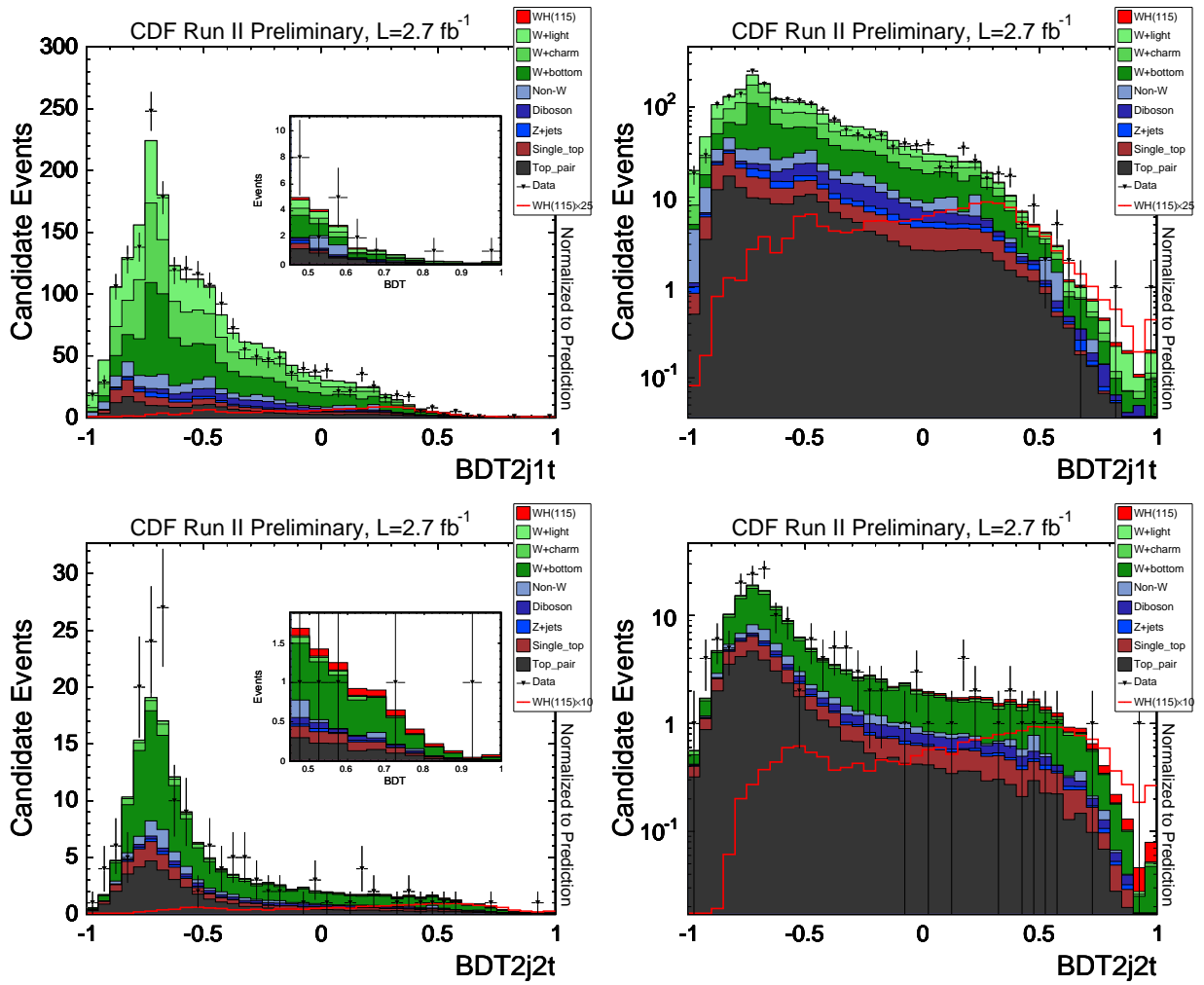


FIG. 4: Top (bottom): BDT output for lepton + 2 jets single (double) tagged data events compared to the Monte Carlo prediction for WH signal ($m_H = 115 \text{ GeV}/c^2$) and background. Left (right) plot shows the BDT output in linear (log) scale.

where $\mathcal{L}^*(data|\beta_{WH})$ is the reduced likelihood and $\pi(\beta_{WH})$ is the prior p.d.f. for β_{WH} . We adopt a flat prior, $\pi(\beta_{WH}) = H(\beta_{WH})$, in this analysis, with H being the Heaviside step function.

To set an upper limit on the WH production cross-section, we integrate the posterior probability density to cover 95% [4]. The observed and expected results are shown in Table V and in Figure 5.

σ / SM	100	105	110	115	120	125	130	135	140	145	150
Expected	3.80	3.98	4.53	5.24	6.34	7.97	9.97	13.4	19.2	27.0	48.7
Observed	3.34	3.56	5.77	6.23	7.08	8.37	11.8	17.1	18.4	44.0	84.1

TABLE V: Expected and observed upper limit cross sections in SM units for different Higgs mass points.

CONCLUSIONS

We have used the matrix element analysis and boosted decision trees technique in a direct search for Higgs boson production in association with a W boson. To extract the most probable WH content in data, we apply a maximum likelihood technique. All sources of systematic rate are included in the likelihood function. We have analyzed

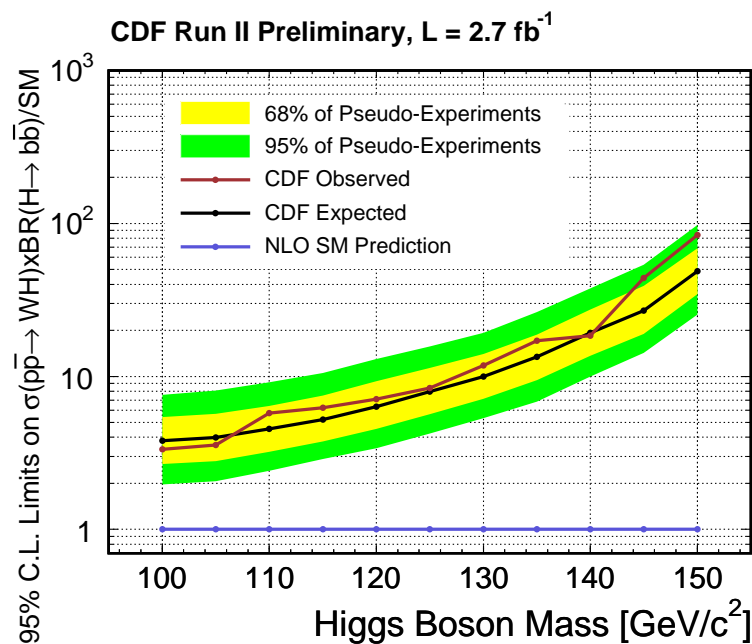


FIG. 5: 95 % C.L. upper limits on the WH production cross-sections times branching ratio for $H \rightarrow b\bar{b}$ for Higgs boson masses between $m_H = 100 \text{ GeV}/c^2$ to $m_H = 150 \text{ GeV}/c^2$. The plot shows the limit normalized to the prediction from the Standard Model.

2.7 fb^{-1} of CDF Run II data. We observe no evidence for a Higgs boson signal and set 95% confidence level upper limits on the WH production cross section times the branching ratio, in SM units, of the Higgs boson to decay to $b\bar{b}$ pairs of $\sigma(p\bar{p} \rightarrow WH) \times BR(H \rightarrow b\bar{b})/SM < 3.34$ to 84.1 for Higgs boson masses between $m_H = 100 \text{ GeV}/c^2$ and $m_H = 150 \text{ GeV}/c^2$. The expected (median) sensitivity estimated in pseudo experiments is $\sigma(p\bar{p} \rightarrow WH) \times BR(H \rightarrow b\bar{b})/SM < 3.80$ to 48.7 at 95% C.L.

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Natural Sciences and Engineering Research Council of Canada; the National Science Council of the Republic of China; the Swiss National Science Foundation; the A.P. Sloan Foundation; the Bundesministerium für Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Science and Technology Facilities Council and the Royal Society, UK; the Institut National de Physique Nucleaire et Physique des Particules/CNRS; the Russian Foundation for Basic Research; the Ministerio de Ciencia e Innovación-Investigación and Programa Consolider-Ingenio 2010, Spain; the Slovak R&D Agency; and the Academy of Finland.

-
- [1] P. Dong, B. Stelzer, R. Wallny, F. Canelli, *Search for Single Top Quark Production using the Matrix Element Technique*, CDF Note 9117 (2008).
 - [2] C. Group, A. Ruiz, B. Casal, R. Vilar, P. Dong, B. Stelzer F. Canelli, *Search for Single Top Quark Production using Boosted Decision Trees*, CDF Note 9117 (2008).
 - [3] F. Caravaglios *et. al.*, M. L. Mangano *et. al.*, JHEP 0307:001 (2003).
 - [4] Particle Data Group, W.-M. Yao *et al.*, J. Phys. **G33**, 1 (2006).
 - [5] A. Hocker, P. Speckmayer, J. Stelzer, F. Tegenfeldt, H. Voss, K. Voss, A. Christov, S. Henrot-Versille, M. Jachowski, A. Krasznahorkay Jr., Y. Mahalalel, R. Ospanov, X. Prudent, M. Wolter, A. Zemla, *TMVA - Toolkit for Multivariate Data Analysis*, arXiv:physics/0703039v4 (2007).
 - [6] L. Demortier, *Bayesian treatments of Systematic Uncertainties*, Proceedings of Advanced Statistical Techniques in Particle Physics, Grey College, Durham, 18 - 22 March 2002.
 - [7] C. Ciobanu, T. Junk, T. Müller, P. Savard, B. Stelzer, W. Wagner, T. Walter, *Likelihood Function for Single Top Search with 162 pb⁻¹*, CDF Note 7106 (2004).
 - [8] L. Read, J.Phys G **28**, 2693 (2002) and T. Junk, Nucl. Instrum. Meth. **434**, 435 (1999). See also P. Bock *et al.* (The LEP

Collaborations), CERN-EP-98-046 (1998) and CERN-EP-2000-055 (2000).

□ $\text{SumE} = \Sigma \frac{pT_{track}}{\sin(2 \cdot \text{atan}(e^{-\eta_{track}}))}$, $\eta = -\frac{1}{2} \cdot \ln(\tan \theta) \Rightarrow \text{SumE} = \Sigma \frac{pT_{track}}{\sin(\theta)}$.

□ 16% of the times there are no jets of cone size 0.7 available. In this case, we use the energy of the jet of cone size 0.4.